

# Consequences of Higher Order Asymptotics for the MSE of M-estimators on Neighborhoods

Peter Ruckdeschel

Received: date / Accepted: date

**Abstract** In Ruckdeschel (2010a), we derive an asymptotic expansion of the maximal mean squared error (MSE) of location M-estimators on suitably thinned out, shrinking gross error neighborhoods. In this paper, we compile several consequences of this result: With the same techniques as used for the MSE, we determine higher order expressions for the risk based on over-/undershooting probabilities as in Huber (1968) and Rieder (1980), respectively. For the MSE problem, we tackle the problem of second order robust optimality: In the symmetric case, we find the second order optimal scores again of Hampel form, but to an  $O(n^{-1/2})$ -smaller clipping height  $c$  than in first order asymptotics. This smaller  $c$  improves MSE only by  $O(n^{-1})$ . For the case of unknown contamination radius we generalize the minimax inefficiency introduced in Rieder et al. (2008) to our second order setup. Among all risk maximizing contaminations we determine a “most innocent” one. This way we quantify the “limits of detectability” in Huber (1997)’s definition for the purposes of robustness.

**Keywords** higher order asymptotics · location M-estimator · second order optimality · minimax radius · cniper contamination

**Mathematics Subject Classification (2000)** MSC 62F12, 62F35

## 1 Motivation/introduction

This paper takes up the central result of Ruckdeschel (2010a): a uniform higher order expansion of the means squared error (MSE) of location M-estimators on suitably

---

P. Ruckdeschel  
Fraunhofer ITWM, Department of Financial Mathematics,  
Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany  
and Dept. of Mathematics, University of Kaiserslautern,  
P.O.Box 3049, 67653 Kaiserslautern, Germany  
E-mail: peter.ruckdeschel@itwm.fraunhofer.de

shrinking and thinned out neighborhoods  $\tilde{Q}_n(r; \varepsilon_0)$ , repeated as Theorem 2.1 in this paper for easier reference. It is of the following form

$$\sup_{Q_n \in \tilde{Q}_n(r; \varepsilon_0)} n \text{MSE}(S_n, Q_n) = r^2 \sup |\psi|^2 + \mathbb{E} \psi^2 + \frac{r}{\sqrt{n}} A_1 + \frac{1}{n} A_2 + o\left(\frac{1}{n}\right) \quad (1.1)$$

Here  $S_n$  is an M-estimator to scores  $\psi$ , and  $A_1, A_2$  are polynomials in the contamination radius  $r$ , in  $b = \sup |\psi|$ , and in the moment functions  $t \mapsto \mathbb{E} \psi_l^t$ ,  $l = 1, \dots, 4$  and their derivatives evaluated in  $t = 0$ , and  $\varepsilon_0$  is the breakdown point of  $S_n$ , i.e.  $\varepsilon_0 = \sup |\psi| / (\sup \psi - \inf \psi)$ . We recognize that the speed of the convergence to the first order as. value is one order faster in the ideal model.

In this paper we present some ramifications of this theorem, but in particular consider its consequences for higher order robust optimality.

**Notation 1.1** For indices we start counting with 0, so that terms of first-order asymptotics have an index 0, second-order ones a 1 and so on. Also we abbreviate first-order, second-order and third-order by f-o, s-o, t-o respectively, and we write f-o-o, s-o-o, and t-o-o for first, second, and third-order asymptotically optimal respectively.

In Theorem 3.1, we take up the over- and undershooting probabilities used as risk in Huber (1968) to determine a finite sample minimax estimator of location. By means of a s-o expansion, we refine the corresponding f-o translation by Rieder (1980), providing a closer link to finite sample optimality.

The closed form expressions in (1.1), in particular under certain symmetry assumptions, allows us to tackle corresponding (uniform) higher order optimality problems, so that we may check whether Pfanzagl (1979)’s catchword “*First order efficiency implies second order efficiency*” survives when passing to neighborhoods around the ideal model, which—at least under symmetry—indeed (partially) holds.

In this setting, we see that Huber-type location M-estimators remain optimal in second order sense, and we even may determine the s-o-o clipping height  $c_1 = c_1(r, n)$  which in fact is slightly lower ( $O(n^{-1/2})$ ) than the f-o-o one. So in fact we only retain the optimal class, not the actual optimal estimator from f-o optimality.

For situations where the radius is (partially) unknown, the concept of a *minimax radius* has been introduced and determined in Rieder et al. (2008): A radius  $r_0$  is determined such that the (f-o) maximal inefficiency  $\bar{\rho}(r')$  (as defined in (5.1)) is minimized in  $r' = r_0$ . We translate this to the s-o setup; the s-o results in the Gaussian location model show that neither  $c_1(r_1, \cdot)$ , nor s-o minimax radius  $r_1(\cdot)$  vary much in  $n$  and that for all  $n$ , s-o minimax inefficiency is always smaller than the corresponding f-o one.

Asymptotics also helps to understand which contaminations are (already) dangerous: We determine the *cniper contamination* as a most innocent appearing least favorable contamination, which is shown to form a saddlepoint together with the f-o (s-o) optimal M-estimator. It appears to be innocent, as it produces only “outliers” which are hardest to detect in some sense specified in this section.

*Organization of the paper* We start with the setup of one dimensional location and recall the main theorem of Ruckdeschel (2010a) in section 2. This result is generalized to a over-/undershooting probability loss in section 3.

Consequences of Theorem 2.1 as to higher order robust optimality are discussed in section 4. As a (partial) explanation for the good, respectively excellent behavior of f-o-o, s-o-o and t-o-o procedures as to numerically exact finite maximal MSE, we present an argument based on a functional implicit function theorem in section 4.2. For decisions upon the procedure to take, only relative risk is relevant which is discussed in some detail in subsection 4.3. Section 5 then considers further supplementary results to Theorem 2.1: a s-o variant of the minimax radius and s-o cniper contaminations. The proofs to the theorems and propositions of this paper are collected in section A.

## 2 Setup

### 2.1 One-dimensional location

We consider estimation of parameter  $\theta$  in a one-dimensional location model, i.e.

$$X_i = \theta + v_i, \quad v_i \stackrel{\text{i.i.d.}}{\sim} F, \quad P_\theta = \mathcal{L}(X_i) \quad (2.1)$$

for some ideal distribution  $F$  with finite Fisher-Information of location  $I(F)$ , i.e.

$$\Lambda_f = -\dot{f}/f \in L_2(F), \quad I(F) = E[\Lambda_f^2] < \infty \quad (2.2)$$

We also assume that  $\Lambda_f$  is increasing. By translation equivariance, we may restrict ourselves to  $\theta_0 = 0$  which will be suppressed in the notation.

The set of *influence curves* (IC's)  $\Psi$  for the estimation of  $\theta$  is defined as Rieder (1994)

$$\Psi := \{\psi \in L_2(F) \mid E[\psi] = 0, \quad E[\psi \Lambda_f] = 1\}, \quad (2.3)$$

where both expectations are evaluated under  $F$ . As class of estimators we consider *asymptotically linear estimators* (ALE's), i.e. estimators  $S_n = S_n(X_1, \dots, X_n)$  with the property

$$\sqrt{n} S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(X_i) + o_{F^n}(n^0) \quad (2.4)$$

We consider maximal mean squared error (MSE) on shrinking neighborhoods of this ideal model, defined as the set  $\mathcal{Q}_n(r)$  of distributions

$$\mathcal{L}_\theta^{\text{real}}(X_1, \dots, X_n) = \mathcal{Q}_n = \bigotimes_{i=1}^n \left[ \left(1 - \frac{r_n}{\sqrt{n}}\right) F + \frac{r_n}{\sqrt{n}} P_{n,i}^{\text{di}} \right] \quad (2.5)$$

with  $r_n = \min(r, \sqrt{n})$ ,  $r > 0$  the contamination radius and  $P_{n,i}^{\text{di}} \in \mathcal{M}_1(\mathbb{B})$  arbitrary, uncontrollable contaminating distributions. As usual, we interpret  $\mathcal{Q}_n$  as the distribution of the vector  $(X_i)_{i \leq n}$  with components

$$X_i := (1 - U_i)X_i^{\text{id}} + U_i X_i^{\text{di}}, \quad i = 1, \dots, n \quad (2.6)$$

for  $X_i^{\text{id}}$ ,  $U_i$ ,  $X_i^{\text{di}}$  stochastically independent,  $X_i^{\text{id}} \stackrel{\text{i.i.d.}}{\sim} F$ ,  $U_i \stackrel{\text{i.i.d.}}{\sim} \text{Bin}(1, r/\sqrt{n})$ , and  $(X_i^{\text{di}}) \sim P_n^{\text{di}}$  for some arbitrary  $P_n^{\text{di}} \in \mathcal{M}_1(\mathbb{B}^n)$ .

Suppressing the dependency upon  $\theta$  as usual, in Rieder (1994), the first order expansion of maximal MSE of an ALE is derived as

$$\tilde{R}(S_n, r) = r^2 \sup |\psi|^2 + E_{\text{id}} |\psi|^2 \quad (2.7)$$

The (first-order) MSE-optimal IC  $\eta_{b_0}$  in a smooth  $p$ -dimensional parametric model with  $L_2$ -derivative  $\Lambda$  by Theorem 5.5.7 (ibid.) has to be of Hampel form

$$\eta_{b_0} = Y \min\{1, b_0/|Y|\}, \quad Y = A\Lambda - a \quad (2.8)$$

for some  $A \in \mathbb{R}^{p \times p}$ ,  $a \in \mathbb{R}^p$  such that  $\eta_{b_0}$  is an IC, and  $b_0$  solving  $E(|Y| - b_0)_+ = r^2 b_0$ . In our location context, for Lagrange multipliers  $z$  and  $A$  such that  $\eta_{b_0} = \eta_{c_0} \in \mathcal{P}$ , we get that

$$\eta_{c_0} = A(\Lambda_f - z) \min\{1, c_0/|\Lambda_f - z|\}, \quad (2.9)$$

$$c_0 \text{ s.t. } E[ (|\Lambda_f - z| - c_0)_+ ] = r^2 c_0 \quad (2.10)$$

## 2.2 Higher Order Expansion

In Ruckdeschel (2010a) we obtain corresponding higher order expansions of the maximal MSE if we thin out the neighborhood system to the set  $\tilde{\mathcal{Q}}_n(r; \varepsilon_0)$  of conditional distributions

$$\mathcal{Q}_n = \mathcal{L}\left\{[(1 - U_i)X_i^{\text{id}} + U_i X_i^{\text{di}}]_i \mid \sum U_i \leq \lceil \varepsilon_0 n \rceil - 1\right\} \quad (2.11)$$

where  $\varepsilon_0 = 1/(2 + \delta_0)$  is the functional (Huber (1981, (2.39),(2.40))) and the finite sample ( $\varepsilon$ -contamination) breakdown point (Donoho and Huber (1983, section 2.2)) of the corresponding M-estimator and  $\delta_0$  is defined by

$$\check{b} := \inf \psi, \quad \hat{b} := \sup \psi, \quad \bar{b} := \frac{1}{2}(\hat{b} - \check{b}), \quad \delta_0 := \frac{|\check{b} + \hat{b}|}{\min((-b), \hat{b})} \geq 0 \quad (2.12)$$

For the result we use the following assumptions and notation: To scores function  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  let  $\psi_t(x) := \psi(x - t)$  and define the following functions  $L(t) := E \psi_t$ ,  $\psi_t^0 := \psi_t - L(t)$ ,  $V(t)^2 := \text{Var} \psi_t$ ,  $\rho(t) := E(\psi_t^0)^3 / V(t)^3$ ,  $\kappa(t) := E[(\psi_t^0)^4] / V(t)^4 - 3$ . Let  $\check{y}_n$  and  $\hat{y}_n$  sequences in  $\mathbb{R}$  such that for some  $\gamma > 1$ ,  $\psi(\check{y}_n) = \inf \psi + o(\frac{1}{n^\gamma})$ ,  $\psi(\hat{y}_n) = \sup \psi + o(\frac{1}{n^\gamma})$ . For  $H \in \mathcal{M}_1(\mathbb{B}^n)$  and an ordered set of indices  $I = (1 \leq i_1 < \dots < i_k \leq n)$  denote  $H_I$  the marginal of  $H$  with respect to  $I$ . Consider three sequences  $c_n$ ,  $d_n$ , and  $\kappa_n$  in  $\mathbb{R}$ , in  $(0, \infty)$ , and in  $\{1, \dots, n\}$ , respectively. We say that the sequence  $(H^{(n)}) \subset \mathcal{M}_1(\mathbb{B}^n)$  is  $\kappa_n$ -concentrated left [right] of  $c_n$  up to  $o(d_n)$ , if for each sequence of ordered sets  $I_n$  of cardinality  $i_n \leq \kappa_n$   $1 - H_{I_n}^{(n)}((-\infty; c_n]^{i_n}) = o(d_n)$ ,  $[1 - H_{I_n}^{(n)}((c_n, \infty)^{i_n}) = o(d_n)]$ . For the theorem we make the following assumptions:

- (bmi)  $\sup \|\psi\| = b < \infty$ ,  $\psi$  monotone,  $\psi \in \mathcal{P}$   
 (D) For some  $\delta \in (0, 1]$ ,  $L$ ,  $V$ ,  $\rho$ , and  $\kappa$  as defined above allow the expansions

$$L(t) = l_1 t + \frac{1}{2} l_2 t^2 + \frac{1}{6} l_3 t^3 + O(t^{3+\delta}), \quad V(t) = v_0(1 + \tilde{v}_1 t + \frac{1}{2} \tilde{v}_2 t^2) + O(t^{2+\delta}) \quad (2.13)$$

$$\rho(t) = \rho_0 + \rho_1 t + O(t^{1+\delta}), \quad \kappa(t) = \kappa_0 + O(t^\delta) \quad (2.14)$$

(Pd) There are some  $T > 0$  and  $\eta > 0$  such that

$$F(t) \geq 1 - t^{-\eta}, \quad \text{for } t > T, \quad F(t) \leq (-t)^{-\eta} \quad \text{for } t < -T \quad (2.15)$$

(C) Let  $f_t$  be the characteristic function of  $\psi_t(X^{\text{id}})$ ; then

$$\lim_{t_0 \rightarrow 0} \limsup_{s \rightarrow \infty} \sup_{|t| \leq t_0} |f_t(s)| < 1 \quad (2.16)$$

With these preparations, we have the following theorem (Ruckdeschel (2010a, Thm. 3.5))

**Theorem 2.1** *In our one-dim. location model assume (bmi) to (C)*

(a) *the maximal MSE of the M-estimator  $S_n$  to scores-function  $\psi$  expands to*

$$R_n(S_n, r, \varepsilon_0) = r^2 b^2 + v_0^2 + \frac{r}{\sqrt{n}} A_1 + \frac{1}{n} A_2 + o(n^{-1}) \quad (2.17)$$

with

$$A_1 = v_0^2 \left( \pm (4 \tilde{v}_1 + 3 l_2) b + 1 \right) + b^2 + [2 b^2 \pm l_2 b^3] r^2 \quad (2.18)$$

$$\begin{aligned} A_2 = & v_0^3 \left( (l_2 + 2 \tilde{v}_1) \rho_0 + \frac{2}{3} \rho_1 \right) + v_0^4 \left( 3 \tilde{v}_2 + \frac{15}{4} l_2^2 + l_3 + 9 \tilde{v}_1^2 + 12 \tilde{v}_1 l_2 \right) + \\ & + [v_0^2 \left( (3 \tilde{v}_2 + 3 \tilde{v}_1^2 + \frac{15}{2} l_2^2 + 2 l_3 + 12 \tilde{v}_1 l_2) b^2 + 1 \pm (8 \tilde{v}_1 + 6 l_2) b \right) + \\ & \pm 3 l_2 b^3 + 5 b^2] r^2 + \left( \left( \frac{5}{4} l_2^2 + \frac{1}{3} l_3 \right) b^4 \pm 3 l_2 b^3 + 3 b^2 \right) r^4 \end{aligned} \quad (2.19)$$

and we are in the  $-[+]$ -case depending on whether (2.20) or (2.21) below applies.

(b) let  $P_n^{\text{di}} := \bigotimes_{i=1}^n P_{n,i}^{\text{di}}$  be contaminating measures for (2.5). Then  $Q_n$  with  $P_n^{\text{di}}$  as contaminating measures generates maximal risk in (2.17) if for  $k_1 > 1$  and  $k_2 > 2 \vee (\frac{3}{2} + \frac{3}{2\delta})$  with  $\delta$  from (Vb) and  $K_1(n) = \lceil k_1 r \sqrt{n} \rceil$  either

$$(P_n^{\text{di}}) \text{ is } K_1(n)\text{-concentrated left of } \check{y}_n - b \sqrt{k_2 \log(n)/n} \text{ up to } o(n^{-1}) \quad (2.20)$$

or

$$(P_n^{\text{di}}) \text{ is } K_1(n)\text{-concentrated right of } \hat{y}_n + b \sqrt{k_2 \log(n)/n} \text{ up to } o(n^{-1}) \quad (2.21)$$

More precisely, if  $\sup \psi < [>] - \inf \psi$ , the maximal MSE is achieved by contaminations according to (2.20) [(2.21)]. In case  $\sup \psi = - \inf \psi$ , (2.20) [(2.21)] applies if

$$\tilde{v}_1 > [<] - \frac{l_2}{4} \left( \frac{b^2}{v_0^2} (r^2 + 3) \left( 1 + \frac{r}{\sqrt{n}} - \frac{2r^2}{n} \right) + 3 \left( 1 - \frac{b^2}{v_0^2} \right) \right) \quad (2.22)$$

If  $\sup \psi = - \inf \psi$  and there is “=” in (2.22), (2.20) and (2.21) generate the same risk up to order  $o(n^{-1})$ .

*Special cases* Let  $Q_n^0$  be any distribution in  $\tilde{Q}_n$  attaining maximal risk in Theorem 2.1. Under symmetry or more specifically if

$$l_2 = v_1 = \rho_0 = 0, \quad (2.23)$$

we obtain as maximal risk in (2.17)

$$\begin{aligned} n E_{Q_n^0}[S_n^2] &= (r^2 b^2 + v_0^2) \left( 1 + \frac{r}{\sqrt{n}} + \frac{r^2}{n} \right) + \frac{r}{\sqrt{n}} (b^2(1+r^2)) + \frac{r^2}{n} (b^2(5+2r^2)) + \frac{\frac{2}{3} v_0^3 \rho_1 + v_0^4 (3\tilde{v}_2 + l_3)}{n} + \\ &+ \frac{(v_0^2 (3\tilde{v}_2 + 2l_3) b^2) r^2 + \frac{1}{3} l_3 b^4 r^4}{n} + o(n^{-1}), \end{aligned} \quad (2.24)$$

while under  $r = 0$  (with or without (2.23)), we get

$$n E_{F^n}[S_n^2] = v_0^2 + \frac{v_0^3 ((l_2 + 2\tilde{v}_1)\rho_0 + \frac{2}{3}\rho_1)}{n} + \frac{v_0^4 (3\tilde{v}_2 + l_3 + \frac{15}{4}l_2^2 + 12\tilde{v}_1 l_2 + 9\tilde{v}_1^2)}{n} + o(n^{-1}) \quad (2.25)$$

respectively, again under (2.23),

$$n E_{F^n}[S_n^2] = v_0^2 + \frac{\frac{2}{3} v_0^3 \rho_1 + v_0^4 (3\tilde{v}_2 + l_3)}{n} + o(n^{-1}). \quad (2.26)$$

### 3 Other loss functions

One easily shows that under similar condition as for Theorem 2.1, we may replace the squared loss function in the MSE by other loss functions  $\ell$  growing atmost at a polynomial rate. In this respect, Theorem 2.1 easily extends to uniform convergence of other risks on  $\tilde{Q}_n$ , e.g. absolute error ( $\ell(x) = |x|$ ),  $L_k$ -error ( $\ell(x) = |x|^k$ ) for  $1 < k < \infty$ , and certain covering probabilities,  $\ell(x) = I_{(\alpha_1, \alpha_2)}(x)$  for some  $\alpha_1 < \alpha_2 \in \mathbb{R}$ .

As an illustration, we consider this last type of loss function, more specifically in the form in which it arises in the finite minimax estimation theory as in Huber (1968) and in which it has been extended to an as. setup by Rieder (1980): The risk is defined as

$$R^b(S_n, r) = \sup_{Q_n \in Q_n(r)} \max \{ Q_n(S_n > \theta + \frac{\alpha_2}{\sqrt{n}}), Q_n(S_n < \theta - \frac{\alpha_1}{\sqrt{n}}) \} \quad (3.1)$$

Fraiman et al. (2001) have taken up a similar setup with conventional confidence intervals to cover bias and variance simultaneously. We work in the setup of Rieder (1980) here and confine ourselves to the higher order terms of order  $n^{-1/2}$ , but of course an extension to terms up to order  $n^{-1}$  as in Theorem 2.1 is feasible. Due to translation equivariance, it is no restriction to consider the case  $\theta = 0$  only. As in Rieder (1980), we work with a possibly asymmetric partition of the interval of given length  $2a/\sqrt{n}$  laid around the estimator: Using the partition

$$2a = \alpha_1 + \alpha_2 = \alpha_1(S_n) + \alpha_2(S_n), \quad (3.2)$$

we minimize the risk according to Rieder (1980, formulas (2.8) and (2.11) in), if with  $\check{b}$ ,  $\hat{b}$ , and  $\bar{b}$  from (2.12) and

$$\alpha_1 = a - \delta, \quad \alpha_2 = a + \delta, \quad \delta = \frac{\epsilon}{2} (\hat{b} + \check{b}) \quad (3.3)$$

If we now account for terms of order  $\frac{1}{\sqrt{n}}$  we minimize the risk if we use the partition

$$2a = \alpha'_1 + \alpha'_2 = \alpha'_1(S_n) + \alpha'_2(S_n), \quad (3.4)$$

with

$$\alpha'_1 = a - \delta - \delta', \quad \alpha'_2 = a + \delta + \delta', \quad (3.5)$$

$\delta' = \delta'_n$  given in the theorem below. To this end, let

$$s_1 := (-a + r\bar{b})/v_0 \quad (3.6)$$

Then, with  $\Phi$  and  $\varphi$  c.d.f. and density of  $N(0, 1)$  and using the notation of Theorem 2.1, we have

**Theorem 3.1** *For the location model (2.1) of finite Fisher information (2.2), assume (bmi), (D') and (C'). Then for sample size  $n$ , the minimal over-/undershooting probability of an M-estimator  $S_n$  for scores-function  $\psi$  in  $Q_n$  obtains eventually in  $n$  as*

$$\begin{aligned} R^k(S_n) &= \sup_{Q_n \in Q_n} \max\{Q_n(S_n \leq -\frac{\alpha'_1}{\sqrt{n}}), Q_n(S_n \geq \frac{\alpha'_2}{\sqrt{n}})\} = \\ &= R_-(S_n, Q_{n;-}^0) = R_+(S_n, Q_{n;+}^0) \end{aligned} \quad (3.7)$$

with  $Q_{n;-}^0$  resp.  $Q_{n;+}^0$  according to (2.20) resp. (2.21) and

$$\begin{aligned} R_-(S_n, Q_{n;-}^0) &= \Phi(s_1) + \frac{1}{\sqrt{n}v_0} \varphi(s_1) \times \\ &\times \left[ \frac{ra}{2} + 2l_2a\delta - as_1\tilde{v}_1v_0 - \frac{r(\check{b}^2 + \hat{b}^2)s_1}{4v_0} + \frac{r^2\bar{b}}{2} \right] + o\left(\frac{1}{\sqrt{n}}\right) \end{aligned} \quad (3.8)$$

and  $\delta' = \delta'_n$  according to

$$\delta' = \frac{1}{\sqrt{n}} \left( -\frac{r\delta}{2v_0} - \frac{l_2}{2v_0} (a^2 + \delta^2) - \tilde{v}_1v_0s_1\delta - \frac{\rho_0}{6} (s_1^2 - 1) + \frac{r\bar{b}\delta s_1}{v_0^2} + \frac{r^2\delta}{2v_0} \right) \quad (3.9)$$

**Remark 3.2** (a) If  $l_2 = \tilde{v}_1 = 0$  and  $\hat{b} = -\check{b}$ , we obtain the same result as (3.8), if we use the expressions  $b_n := \text{Bias}_n$  and  $v_n^2 = \text{Var}_n$  for bias and variance from Ruckdeschel (2010a, Prop.6.4), plug them into the as. risk, which gives  $\Phi((rb_n - a)/v_n)$ , and then expand this up to  $o(n^{-1/2})$ .

(b) The numerical values obtainable by Theorem 3.1 should be compared to those of Kohl (2005, sections 11.3.3.3 and 11.4.1); admittedly the approach of Theorem 3.1 in this context gives rather poor (too liberal) approximations compared to those in the cited reference (see the R-file `Thm31.R` available on the web-page to this article); this is plausible though, as Kohl already starts with finitely optimal procedures whereas our approach improves upon asymptotically optimal ones.

## 4 Consequences: Higher Order Optimality and Relative Risk

In this section, we consider the class  $\mathcal{S}_2$  of all M-estimators according to (bmi), (D'), and (C') as well as (Pd); correspondingly, we define  $\mathcal{S}_3$  with (D), (C) replacing (D'), (C'); we always assume that the class of M-estimators  $\mathcal{H}$  of ICs of Hampel-type (2.9) forms a subset of  $\mathcal{S}_2$  [ $\mathcal{S}_3$ ]. In particular we assume  $f$  to be log-concave.

#### 4.1 Second-order optimality

Symmetry allows considerable simplifications; for instance, if  $F$  is symmetric, i.e.  $F(B) = F(-B)$  for all  $B \in \mathbb{B}$ , in (2.9) always  $z = 0$ . But also, much deeper results are possible. Thus for the rest of this subsection, we assume (2.23). Then (2.24) gives the s-o-maximal MSE for any M-estimator in  $\mathcal{S}_2$ ; in particular

$$A_1 = v_0^2 + b^2(1 + 2r^2) \quad (4.1)$$

Condition (2.23) clearly holds for skew symmetric  $\psi$  and symmetric  $F$ . For symmetric  $F$ , however, for any IC  $\psi$ , also  $\tilde{\psi} := -\psi(-\cdot)$  is an IC and hence so is the skew-symmetrized  $\psi^{(s)} := \frac{1}{2}(\psi + \tilde{\psi})$ , too. But by convexity of the MSE,  $\psi^{(s)}$  will be at least as good as  $\psi$  as to MSE, hence it is no restriction to only consider skew symmetric ICs, and we fall into the application range of Ruckdeschel and Rieder (2004, Thm. 3.1), i.e.,

**Theorem 4.1** *Assume that maximal as. risk of an ALE on  $\tilde{\mathcal{Q}}_n$  resp.  $\tilde{\mathcal{Q}}'_n(\cdot, s_0)$  is representable as  $G(rb(\psi), v_0(\psi))$  for some convex real-valued function  $G(w, s)$ , strictly isotone in both arguments and totally differentiable, bounded away from the minimum for  $w \rightarrow \infty$ . Then, on  $\mathcal{Q}_n$ , respectively on  $\tilde{\mathcal{Q}}_n$ , the optimal IC of Hampel-type (2.9) for some clipping height  $b = Ac$  determined by*

$$r v_0 \partial_w G(rAc, v_0) = \partial_s G(rAc, v_0) A E(|\Lambda| - c)_+ \quad (4.2)$$

In our case, this theorem specializes to

**Corollary 4.2** *Assume a symmetric model (2.1) with increasing  $\Lambda_f$  and (2.2). Under the assumptions of this section, the s-o-o M-estimator in class  $\mathcal{S}_2$  has an IC of Hampel-type (2.9) with  $z = 0$  and the s-o-o clipping height  $c_1 = c_1(n)$  is determined by*

$$r^2 c \left( 1 + \frac{r^2 + 1}{r^2 + r\sqrt{n}} \right) = E(|\Lambda| - c)_+ \quad (4.3)$$

Always,  $c_0 > c_1(n)$ . Suppose that  $h(c) := E(|\Lambda| - c)_+$  is differentiable in  $c_0$  with derivative  $h'(c_0)$ . Then,

$$c_1(n) = c_0 \left( 1 - \frac{1}{\sqrt{n}} \frac{r^3 + r}{r^2 - h'(c_0)} \right) + o\left(\frac{1}{\sqrt{n}}\right) \quad (4.4)$$

That is, (for  $n$  large enough) **the f-o-o clipping height  $c_0$  always is too optimistic.**

Assume s-o risk of ICs of Hampel-type (2.9) is smooth enough in  $c$  in its minimum  $c_1$  to allow a s-o Taylor expansion, which is an assumption on the remainder  $o(n^{-1})$  present in (2.17). Then, around  $c_1$ , s-o risk behaves like a parabola. But, as by (4.4),  $c_1 - c_0 = O(1/\sqrt{n})$ , using  $c_1$  instead of  $c_0$  can only improve s-o risk by order  $O(1/n)$ . This even carries over to risks “near” s-o risk:



## 4.2 Consequences for the exact MSE

**Proposition 4.3** *Let  $F, F_n, G_n \in C_2(\mathbb{R})$ ,  $n \in \mathbb{N}$ , such that for some  $\beta \geq \beta' > 0$*

$$\begin{aligned} \text{(i)} \quad & \sup_x |F_n - G_n| + |F'_n - G'_n| + |F''_n - G''_n| = O(n^{-\beta}), \\ \text{(ii)} \quad & \sup_x |F_n - F| + |F'_n - F'| + |F''_n - F''| = O(n^{-\beta'}) \end{aligned} \quad (4.5)$$

*Assume that in  $x_0 \in \mathbb{R}$ ,  $F(x_0)$  is minimal, and that  $F''(x_0) = f_2 > 0$ . Then*

- (a) there is some sequence  $(x_n) \subset \mathbb{R}$  such that eventually in  $n$ ,  $F_n(x_n)$  is minimal and  $\lim F''_n(x_n) = f_2$ .*
- (b)  $|x_n - x_0| = O(n^{-\beta'})$ .*
- (c) there is some sequence  $(y_n) \subset \mathbb{R}$  such that eventually in  $n$ ,  $G_n(y_n)$  is minimal and  $\lim G''_n(y_n) = f_2$ .*
- (d)  $|y_n - x_0| = O(n^{-\beta})$ .*
- (e)  $0 \leq G_n(x_n) - G_n(y_n) = O(n^{-2\beta})$ .*

The drawback of this proposition is that assumption (4.5) is difficult to check if we have no explicit expression for  $G_n$ : For given  $r \geq 0$ , let  $\text{asMSE}_{j=0,1,2}(c)$  be the f-o, s-o, and t-o maximal MSE of an M-estimator in  $\mathcal{H}$ , and  $\text{exMSE}(c)$  the corresponding exact maximal MSE  $R_n$ ; we would like to apply Proposition 4.3 to  $F = \text{asMSE}_0$ ,  $F_n = \text{asMSE}_{j=1,2}$  and  $G_n = \text{exMSE}$  to conclude on the performance of f-o-o, s-o-o, t-o-o procedures as to  $\text{exMSE}$ . As to (4.5), part (ii) is easy to see checking the expressions, giving  $\beta' = 1/2$ , while for part (i) Theorem 2.1 only says that  $\sup_x |F_n - G_n| = o(n^{-j/2})$  which in fact is  $O(n^{-(j/2+\delta)})$ , and probably, under slightly stronger assumptions,  $O(n^{-(j+1)/2})$ . So presumably—in view of Table 2,

$$0 \leq \text{exMSE}(c_{j,n}) - \text{exMSE}(c_{\text{ex};n}) = O(n^{-j-1}), \quad j = 0, 1, 2 \quad (4.6)$$

**Remark 4.4** We even conjecture that we may apply an analogue to Proposition 4.3 for functions  $F, F_n, G_n: \mathcal{P} \rightarrow \mathbb{R}$ : Let us denote by  $\hat{\psi}^{(j;n)}$ , the corresponding f-o, s-o, t-o optimal IC and  $\hat{\psi}^{(\text{ex};n)}$  the exactly optimal IC; then, with the usual abuse of notation as to  $\text{exMSE}$ , we conjecture that

$$0 \leq \text{exMSE}(\hat{\psi}^{(j;n)}) - \text{exMSE}(\hat{\psi}^{(\text{ex};n)}) = O(n^{-j-1}), \quad j = 0, 1, 2 \quad (4.7)$$

## 4.3 Relative risk

An observation in the simulation study was that the relative MSE w.r.t. the MSE of the f-o-o procedure seemed to converge faster than the absolute terms. This is reflected by our formulas as follows:

### 4.3.1 Contaminated situation

Let  $\text{asMSE}_0(c)$  and  $A_1(c)$  be the f-o as. MSE and the corresponding s-o correction term for the Hampel-IC with clipping height  $c$ . Then we may write for the f-o [s-o]

relative risk  $\text{relMSE}_0(c, r)$  [ $\text{relMSE}_1(c, r, n)$ ] w.r.t. the corresponding risk of the f-o-o procedure

$$\begin{aligned} \text{relMSE}_1(c, r, n) &:= \frac{\text{asMSE}_0(c) + \frac{r}{\sqrt{n}} A_1(c)}{\text{asMSE}_0(c_0) + \frac{r}{\sqrt{n}} A_1(c_0)} = \\ &= \text{relMSE}_0(c, r) \left( 1 + \frac{r}{\sqrt{n}} (\Delta(c) - \Delta(c_0)) \right) + o(n^{-1/2}) \end{aligned} \quad (4.8)$$

with

$$\Delta(c) := \frac{b^2(c) - v_0^2(c)}{\text{asMSE}_0(c)} \quad (4.9)$$

So in fact, the observed faster convergence is not reflected by higher order optimality, but as we will see, the difference between  $\text{relMSE}_0(c, r)$  and  $\text{relMSE}_1(c, r)$  are in fact small.

Procedure choice will usually be based on relative risk, so it is interesting to consider the maximal error compared to the s-o approximation one incurs when using the f-o asymptotics instead. In view of subsection 4.1 we will limit ourselves to only considering Hampel-IC's with a clipping height  $c$  in the range

$$C(c_0, \rho) := [c_0/(1 + \rho), c_0(1 + \rho)], \quad (4.10)$$

for  $\rho \geq 0$ . This leads us to

$$\Delta \widehat{\text{relMSE}}(r; \rho) := \max_{c \in C(c_0(r), \rho)} r (\Delta(c) - \Delta(c_0(r))) \quad (4.11)$$

or even maximizing over the radius

$$\widehat{\Delta}(\rho) := \Delta \widehat{\text{relMSE}}(\rho) := \max_r \Delta \widehat{\text{relMSE}}(r; \rho) \quad (4.12)$$

In the Gaussian case, the function  $r \mapsto \Delta \widehat{\text{relMSE}}(r; \rho)$  is plotted for  $\rho = 0.1$  in Figure 1, and for  $\Delta(0.1)$ , we get a value of 0.065, which for an actual sample size  $n$  has to be divided by  $\sqrt{n}$ —an astonishingly good approximation!

**So down to very moderate sample sizes we can base our decision which clipping height to take to achieve “nearly” the optimal MSE on  $\tilde{Q}_n$  on f-o asymptotics only.** A similar consideration is of course possible for the ideal situation.

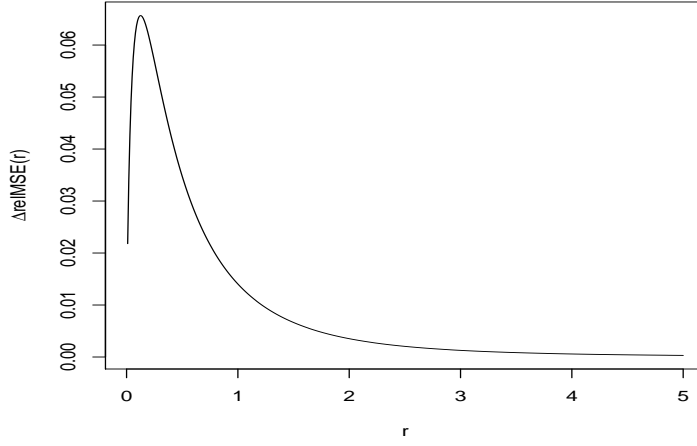
#### 4.3.2 Illustration

As an example we take  $F = \mathcal{N}(0, 1)$  and calculate the terms  $c_1$ ,

$$\text{asMSE}_1 := \text{asMSE}_0 + \frac{r}{\sqrt{n}} A_1 \quad (4.13)$$

and  $\text{relMSE}_1$  for the radii and sample sizes of the simulation study where for the optimization for  $c_1$  we use the function `optimize` in R 2.11.0 (compare R Development Core Team (2010)). The results are tabulated in Table 1. Correspondingly, we also determine the t-o terms  $c_2$ ,

$$\text{asMSE}_2 := \text{asMSE}_1 + A_2/n \quad (4.14)$$



**Fig. 1** The mapping  $r \mapsto \Delta \widehat{\text{relMSE}}(r; \rho)$  for  $F = \mathcal{N}(0, 1)$  and for  $\rho = 0.1$ .

**Table 1**  $c_1(r, n)$ ,  $\text{asMSE}_1(c_1(r, n), r, n)$  and  $\text{relMSE}_1(c_1(r, n), r, n)$

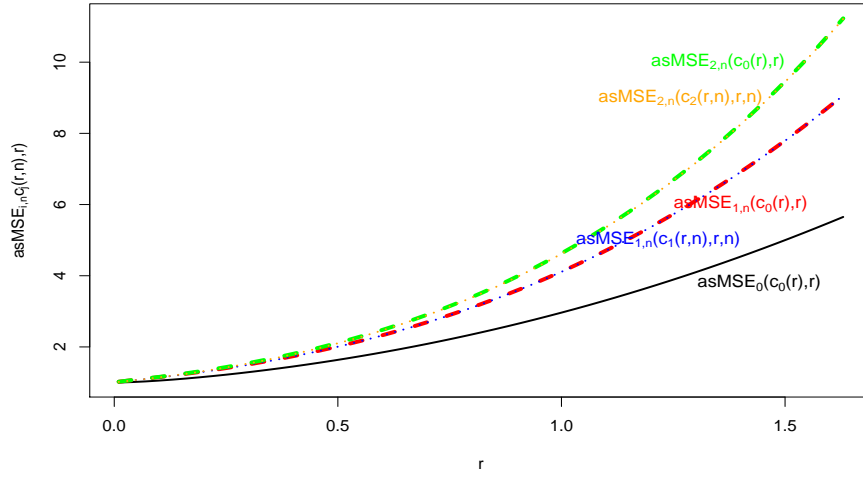
$r$		$n = 5$	$n = 10$	$n = 30$	$n = 50$	$n = 100$	$n = \infty$
0.1	$c_1$	1.394	1.484	1.611	1.663	1.724	1.948
	$\text{asMSE}_1$	1.248	1.197	1.140	1.122	1.103	1.054
	$\text{relMSE}_1$	3.476%	2.149%	0.939%	0.623%	0.349%	0.000%
0.25	$c_1$	0.994	1.059	1.147	1.181	1.219	1.339
	$\text{asMSE}_1$	1.635	1.519	1.397	1.358	1.319	1.220
	$\text{relMSE}_1$	2.377%	1.470%	0.632%	0.414%	0.228%	0.000%
0.5	$c_1$	0.650	0.690	0.746	0.767	0.790	0.862
	$\text{asMSE}_1$	2.527	2.271	2.006	1.923	1.840	1.636
	$\text{relMSE}_1$	1.214%	0.772%	0.342%	0.226%	0.126%	0.000%
1.0	$c_1$	0.320	0.340	0.369	0.380	0.394	0.436
	$\text{asMSE}_1$	5.761	4.944	4.110	3.852	3.593	2.964
	$\text{relMSE}_1$	0.427%	0.292%	0.142%	0.098%	0.056%	0.000%

and in Figure 2, we plot the graphs of the five functions

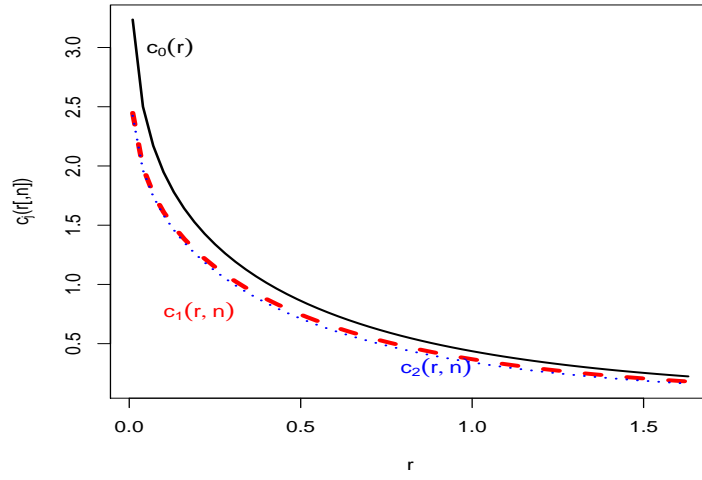
$$\begin{aligned}
 r &\mapsto \text{asMSE}_0(\eta_{c_0(r)}, r), & r &\mapsto \text{asMSE}_1(\eta_{c_0(r)}, r, n), & r &\mapsto \text{asMSE}_2(\eta_{c_0(r)}, r, n) \\
 r &\mapsto \text{asMSE}_1(\eta_{c_1(r, n)}, r, n), & r &\mapsto \text{asMSE}_2(\eta_{c_2(r, n)}, r, n)
 \end{aligned}$$

for  $F = \mathcal{N}(0, 1)$  and for  $n = 30$ . In fact, the choice of the clipping height— $c_0(r)$ ,  $c_1(r, n)$ ,  $c_2(r, n)$ —does not entail any visible changes while the absolute value of f-o, s-o, and t-o MSE clearly differ.

In the same situation, the three functions  $r \mapsto c_0(r)$ ,  $r \mapsto c_1(r, n)$ ,  $r \mapsto c_2(r, n)$  are plotted in Figure 3; while there are visible differences between  $c_0(r)$  and  $c_i(r, n)$ ,  $i = 1, 2$ ,  $c_1(r, n)$  and  $c_2(r, n)$  visually coincide.



**Fig. 2** The mapping  $r \mapsto \text{asMSE}_{i,n}(\eta_{c_j(r,n)}, r, n]$  for  $i = 0, 1, 2$ ,  $j = 0, i$ ,  $n = 30$  and  $F = \mathcal{N}(0, 1)$



**Fig. 3** The mapping  $r \mapsto c_j(r, n]$  for  $j = 0, 1, 2$  and  $n = 30$  and  $F = \mathcal{N}(0, 1)$

#### 4.4 Comparison with the approach by Fraiman et al. (2001)

Fraiman et al. (2001) work in a similar setup, i.e. the one-dimensional location problem where the center distribution is  $F_0 = \mathcal{N}(0, \sigma^2)$  and an M-estimator  $S_n$  to skew symmetric scores  $\psi$  is searched which minimizes the maximal risk on a neighborhood about  $F_0$ . Contrary to our approach, the authors work with convex contamination neighborhoods  $\mathcal{V} = \mathcal{V}(F, \varepsilon)$  to a fixed radius  $\varepsilon$ .

There has been some discussion which approach—fixed or shrinking radius—is more appropriate, but for fixed sample size  $n$ , of course we may translate the fixed radius  $\varepsilon$  into our radius  $r/\sqrt{n}$  and then compare the approximation quality of both approaches. Fraiman et al. (2001) propose to use risks which are constructed by means of a positive function  $g : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$  of as. bias  $B = B(F, \psi)$  and as. variance  $v^2 = V^2(F, \psi)$ . Here,  $B$  is defined as zero of  $\beta \mapsto (1 - \varepsilon) \int \psi_\beta dF + \varepsilon b$ , and  $v^2 := V_1/V_2^2$  for  $V_1 = (1 - \varepsilon) \int \psi_B^2 dF + \varepsilon b^2$  and  $V_2 = (1 - \varepsilon) \int \psi_B dF$ .

Function  $g$  is assumed lower semicontinuous and symmetric in the first argument as well as isotone in each argument. The risk of an M-estimator to IC  $\psi$  is taken as the function

$$L_g(\psi) = \sup_{G \in \mathcal{V}} g(B(G, \psi), v(G, \psi)/n) \quad (4.15)$$

A MSE-type risk then is given by  $g(u, v) = u^2 + v$ . It is not quite MSE, as it employs the as. terms  $B$  and  $v$ , so their results may differ from ours. The crucial point is that to solve their optimization problem, the authors have to assume that besides bias, also variance is maximized (for their optimal  $\hat{\psi}$ ) if we contaminate with a Dirac measure in  $\infty$ . According to this assumption, if we introduce  $G_0 := (1 - \varepsilon)F_0 + \varepsilon I_{\{\infty\}}$ , we have to find  $\psi$  minimizing

$$l_g(\psi) = g(B(G_0, \psi), v(G_0, \psi)/n) \quad (4.16)$$

Differently to the Hampel-type IC's the solutions to this problem are of form

$$\psi_{a,b,c,t}(x) = \tilde{\psi}_{a,b,t}(x \min\{1, \frac{c}{|x|}\}), \quad (4.17)$$

$$\tilde{\psi}_{a,b,t}(x) = a \tanh(tx) + b[x - t \tanh(tx)] \quad (4.18)$$

but the “MSE”-optimal solutions are numerically quite close to corresponding Hampel-ICs  $\psi_H$ , for which the authors in turn show that always  $L_g(\psi_H) = l_g(\psi_H)$ .

For an implementation of this optimization see the R-file `FYZ.R` available on the webpage.

#### A comparison

As a sort of benchmark for our results, we reproduce a comparison to be found in Ruckdeschel and Kohl (2010)—albeit in some more detail than in the cited reference: For a set of values for  $n$  and  $r$ , we determine the “MSE”-optimal  $\hat{\psi}$  and a corresponding Hampel IC  $\hat{\psi}_H$  which is then compared to the f-o-o and s-o-o IC derived in this paper. Within the class of Hampel-IC's, numerically, we also determine the t-o-o and the “exactly” optimal clipping- $c$ ,  $c_2$  and  $c_{ex}$  respectively. We compare the resulting IC's as to their clipping-height and the corresponding (numerically exact) value of

$R_n(S_n, r)$ , denoted by  $\text{MSE}_n$ ; the latter comparison is done by the terms  $\text{relMSE}_n^{\text{ex}}(c.)$ , calculated as

$$\text{relMSE}_n^{\text{ex}}(c.) = (\text{MSE}_n(c.) / \text{MSE}_n(c_{\text{ex}}) - 1) \times 100\% \quad (4.19)$$

The results are displayed in Table 2. Also compare the function `allMSEs` in the R-file `asMSE.R` available on the web-page to this article.

For the numerical evaluation of the MSE, we use Algorithms C (more accurate, but slow for larger  $n$ ) and D (a little inaccurate for small  $n$ , but fast) discussed in Ruckdeschel and Kohl (2010). For  $n = \infty$ , we evaluate the corresponding f-o as. MSE for the IC to the corresponding values of  $c$ . As a cross-check, the clipping heights  $c_i$ ,  $i = 0, 1, 2$  are also determined for  $n = 10^8$ . In case of  $c_{\text{FZY}}$ , for all finite  $n$ 's the error tolerance used in `optimize` in R was  $10^{-4}$ , while for  $n = \infty$  it was  $10^{-12}$ . For  $c_{\text{ex}}$  and  $n = 10^8$ , an optimization of the (numerically) exact MSE would have been too time-consuming and has been skipped for this reason. Also, for  $n = 5$ , the radius  $r = 1.0$ , corresponding to  $\varepsilon = 0.447$ , is not admitted for an optimization of (4.16) and thus no result is available in this case.

## 5 Ramifications: Minimax radius and Cniper contamination

### 5.1 Minimax radius

In this subsection, we refine the results of Rieder et al. (2008). In the cited paper, we want to give a guideline to the statistician which procedure to choose if he knows that there is contamination but does not know the radius exactly: To this end, we consider the maximal inefficiency  $\bar{\rho}(r')$  defined as

$$\bar{\rho}_0(r') := \sup_{r \in (r_l, r_u)} \bar{\rho}(r', r), \quad \bar{\rho}(r', r) := \frac{\bar{R}(\eta_{c_0(r')}, r)}{\bar{R}(\eta_{c_0(r)}, r)} \quad (5.1)$$

and determine the minimax radius  $r_0$  as minimizer of  $\bar{\rho}_0(r')$ . If one knows at least that the actual radius will lie in an interval  $[r/\gamma, r\gamma]$  we may determine  $r_{\gamma,r}$  as minimizer of  $\bar{\rho}_\gamma(r', r) = \sup_{s \in (r/\gamma, r\gamma)} \bar{\rho}(r', s)$  and denote the corresponding minimax inefficiency by  $\bar{\rho}_\gamma(r)$ . In a second optimizing step we then determine the maximizer  $r_\gamma$  of  $\bar{\rho}_\gamma(r)$ . The unrestricted case is symbolically included by  $\gamma = \infty$ . In the Gaussian location case this gives

$\gamma = 0$			$\gamma = 2$			$\gamma = 3$		
$r_0$	$c_0(r_0)$	$\bar{\rho}_0(r_0)$	$r_2$	$c_0(r_2)$	$\bar{\rho}_2(r_2)$	$r_3$	$c_0(r_3)$	$\bar{\rho}_3(r_3)$
0.621	0.718	18.07%	0.575	0.769	8.84%	0.549	0.799	4.41%

These calculations can easily be translated to the s-o setup setting

$$R_1(\psi, r, n) := r^2 \sup |\psi|^2 + \mathbb{E} \psi^2 + \frac{r}{\sqrt{n}} A_1 \quad (5.2)$$

so that in this paper we would instead determine  $r_1(n)$  as minimizer of  $\rho_1(r', r, n)$ ,

$$\sup_{r \in (r_l, r_u)} \rho_1(r', r, n), \quad \rho_1(r', r, n) := \frac{R_1(\eta_{c_1(r'(n), n)}, r, n)}{R_1(\eta_{c_1(r, n)}, r, n)} \quad (5.3)$$

**Table 2** Optimal clipping heights and corresponding (numerically) exact MSE

$r$		$n = 5$	$n = 10$	$n = 30$	$n = 50$	$n = 100$	$n = \infty$
0.1	$c_0$	1.948	1.948	1.948	1.948	1.948	1.948
	$\text{relMSE}_n^{\text{ex}}(c_0)$	8.679%	4.065%	1.340%	0.836%	0.448%	—
	$c_1$	1.394	1.484	1.611	1.663	1.724	1.948
	$\text{relMSE}_n^{\text{ex}}(c_1)$	0.833%	0.207%	0.027%	0.014%	0.010%	—
	$c_2$	1.309	1.428	1.585	1.644	1.713	1.948
	$\text{relMSE}_n^{\text{ex}}(c_2)$	0.332%	0.066%	0.008%	0.004%	0.006%	—
	$c_{\text{FZY}}$	1.368	1.370	1.610	1.668	1.756	1.939
	$\text{relMSE}_n^{\text{ex}}(c_{\text{FZY}})$	0.658%	0.002%	0.026%	0.021%	0.031%	—
0.25	$c_{\text{ex}}$	1.167	1.358	1.560	1.630	1.704	—
	$\text{MSE}_n(c_{\text{ex}})$	1.388	1.239	1.151	1.129	1.107	—
	$c_0$	1.339	1.339	1.339	1.339	1.339	1.339
	$\text{relMSE}_n^{\text{ex}}(c_0)$	6.280%	3.681%	1.108%	0.656%	0.330%	—
	$c_1$	0.994	1.059	1.147	1.181	1.219	1.339
	$\text{relMSE}_n^{\text{ex}}(c_1)$	0.933%	0.415%	0.055%	0.023%	0.009%	—
	$c_2$	0.890	0.990	1.114	1.159	1.207	1.339
	$\text{relMSE}_n^{\text{ex}}(c_2)$	0.241%	0.104%	0.009%	0.002%	0.003%	—
0.5	$c_{\text{FZY}}$	0.924	1.020	1.205	1.177	1.211	1.338
	$\text{relMSE}_n^{\text{ex}}(c_{\text{FZY}})$	0.417%	0.215%	0.233%	0.018%	0.002%	—
	$c_{\text{ex}}$	0.783	0.921	1.092	1.140	1.205	—
	$\text{MSE}_n(c_{\text{ex}})$	2.225	1.705	1.438	1.381	1.330	—
	$c_0$	0.862	0.862	0.862	0.862	0.862	0.862
	$\text{relMSE}_n^{\text{ex}}(c_0)$	2.930%	2.655%	0.792%	0.446%	0.218%	—
	$c_1$	0.650	0.690	0.746	0.767	0.790	0.862
	$\text{relMSE}_n^{\text{ex}}(c_1)$	0.756%	0.615%	0.087%	0.036%	0.013%	—
1.0	$c_2$	0.547	0.620	0.712	0.744	0.777	0.862
	$\text{relMSE}_n^{\text{ex}}(c_2)$	0.230%	0.191%	0.015%	0.008%	0.003%	—
	$c_{\text{FZY}}$	0.539	0.632	0.716	0.749	0.782	0.866
	$\text{relMSE}_n^{\text{ex}}(c_{\text{FZY}})$	0.200%	0.248%	0.021%	0.011%	0.008%	—
	$c_{\text{ex}}$	0.413	0.531	0.686	0.728	0.770	—
	$\text{MSE}_n(c_{\text{ex}})$	4.632	3.039	2.162	2.008	1.879	—
	$c_0$	0.436	0.436	0.436	0.436	0.436	0.436
	$\text{relMSE}_n^{\text{ex}}(c_0)$	2.716%	3.132%	0.746%	0.348%	0.149%	—
1.0	$c_1$	0.320	0.340	0.369	0.380	0.394	0.436
	$\text{relMSE}_n^{\text{ex}}(c_1)$	1.411%	1.610%	0.251%	0.076%	0.021%	—
	$c_2$	0.255	0.291	0.342	0.361	0.382	0.436
	$\text{relMSE}_n^{\text{ex}}(c_2)$	0.876%	0.999%	0.123%	0.027%	0.006%	—
	$c_{\text{FZY}}$	—	0.281	0.344	0.375	0.387	0.440
	$\text{relMSE}_n^{\text{ex}}(c_{\text{FZY}})$	—	0.892%	0.132%	0.063%	0.012%	—
	$c_{\text{ex}}$	0.001	0.125	0.286	0.334	0.366	—
	$\text{MSE}_n(c_{\text{ex}})$	12.627	8.445	4.948	4.296	3.787	—

$c$	order	determined by	optimal among M-estimators	
$c_0$	f-o-o	num. solution of (2.10)	to any IC	where
$c_1$	s-o-o	num. solution of (4.3)	in $\mathcal{S}_2$ (see section 4.1)	
$c_2$	t-o-o	num. optimization of (2.17)	in $\mathcal{H}$ (see section 4.1)	
$c_{\text{FZY}}$	—	num. optimization of (4.16)	to (4.18)-type ICs	
$c_{\text{ex}}$	—	num. optimization of the (num.) exact MSE	in $\mathcal{H}$ (see section 4.1)	

(4.3) is the s-o analogue to (2.10), which is derived in Corollary 4.2. A more detailed description to this table is located on page 13.

Table 3: Minimax radii for second order asymptotics

		$n = 5$	$n = 10$	$n = 30$	$n = 50$	$n = 100$	$n = \infty$
$\gamma = 0$	$r_\gamma$	0.390	0.449	0.514	0.536	0.559	0.621
	$c_1(r_\gamma)$	0.776	0.749	0.729	0.725	0.722	0.718
	$\rho_{1;\gamma}(r_\gamma)$	16.27%	17.08%	17.71%	17.85%	17.96%	18.07%
$\gamma = 3$	$r_\gamma$	0.481	0.496	0.518	0.524	0.534	0.548
	$c_1(r_\gamma)$	0.670	0.694	0.724	0.739	0.750	0.800
	$\rho_{1;\gamma}(r_\gamma)$	6.213%	6.773%	7.490%	7.751%	8.036%	8.836%
$\gamma = 2$	$r_\gamma$	0.540	0.552	0.564	0.563	0.571	0.574
	$c_1(r_\gamma)$	0.609	0.637	0.675	0.695	0.707	0.770
	$\rho_{1;\gamma}(r_\gamma)$	2.987%	3.297%	3.692%	3.834%	3.988%	4.410%

respectively  $\rho_{1;\gamma}$  and instead of  $\bar{\rho}_\gamma$ . For finite  $n$ , however, we have to take into account that  $r < \sqrt{n}$  always. Doing so we get Table 3, showing that there is not much variation in both  $c_1(r_\infty, \cdot)$ ,  $\rho_{1;\gamma}(r_\gamma, \cdot)$  for varying  $n$ .

**So if  $r$  is completely unknown, it is a good choice to use the M-estimator to Hampel-scores for  $c \approx 0.7$ —you will never have a larger inefficiency than the limiting 18%! Ex post this is one more argument, why the H07-estimate survived in in Sections 7.B.8 and 7.C.4 of the Princeton robustness study (Andrews et al. (1972)).** A table for the corresponding t-o minimax radii is available on the web-page.

## 5.2 Innocent-looking risk-maximizing contaminations

In Huber (1997, p. 62), the author complains “... the considerable confusion between the respective roles of diagnostics and robustness. The purpose of robustness is to safeguard against deviations from the assumptions, in particular against those that are near or below the limits of detectability.” As worked out in Ruckdeschel (2006), the exact critical rate for these limits may be determined in a statistical way: For some prescribed outlier set OUT, let  $p_0$  and  $q_n = (1 - r_n)p_0 + r_n$  be the probability under the ideal model, and under convex contaminations of radius  $r_n$ , respectively. Considering the minimax test between these alternatives yields the exact critical rate  $1/\sqrt{n}$ : under a faster shrinking  $p_0$  cannot be separated from  $q_n$  at all, while at a slower rate, asymptotically we can separate them without error.

Going one step further, for some given  $1/\sqrt{n}$ -shrinking neighborhoods of radius  $r$ , we would also like to know how “small” an outlier may be, while it is still harmful enough to distort the classically optimal procedure in a way that this procedure is beaten by some robust one.

### 5.2.1 The Cniper contaminaton

To a fixed radius  $r$ , in the preceding sections, we have found/discussed f-o-o and s-o-o ICs of Hampel-form with clipping height  $c_j = c_j(r, n)$ ,  $j = 0, 1$ . To these ICs we have derived families of contaminations achieving maximal risk on  $\tilde{Q}_n(r)$ . By means of Theorem 2.1(b), these are induced by any contaminating measures  $P_n^{\text{di}}$  under which



$\eta_\theta(X^{\text{di}})$  is constantly either  $b_j$  or  $-b_j$  for  $b_j = A_j c_j$ —up to an event of probability  $o(n^{-1})$ . Out of these risk-maximizing contaminations, let us limit ourselves to those induced by Dirac masses at  $x$ :

$$Q_n(x) := [(1 - \frac{r}{\sqrt{n}})P_\theta + \frac{r}{\sqrt{n}} \mathbf{I}_{\{x\}}]^{\otimes n} \quad (5.4)$$

Among these  $Q_n(x)$ , we seek the least “suspicious” looking contamination point  $x$  in the sense that the region  $\text{OUT}_j := [x; \infty)$  [or  $(-\infty; x]$ ] carries large ideal probability. With this region as outlier set in Ruckdeschel (2006), values of  $x$  (or slightly above in absolute value) occurring more frequently than they should under the ideal situation, are hardest to detect.

More precisely, in the general smooth parametric setup (compare Kohl et al. (2010)), assume that the observations are univariate; let  $S_n^{(b_0)}$  and  $\hat{S}_n$  be ALEs to the classical optimal IC  $\hat{\eta} = \mathcal{I}^{-1}\Lambda$  and the asMSE<sub>0</sub>-optimal IC  $\eta_{b_0}$ , respectively. In this setup we define

**Definition 5.1** *The f-o cniper point  $x_0$  is defined as  $x_{0,+}$  if  $x_{0,+} \geq -x_{0,-}$  and  $x_{0,-}$  else, where*

$$\begin{aligned} x_{0,+} &:= \inf\{x > 0 \mid \text{asMSE}_0(S_n^{(b_0)}, Q_n(x)) < \text{asMSE}_0(\hat{S}_n, Q_n(x))\} \\ x_{0,-} &:= \sup\{x < 0 \mid \text{asMSE}_0(S_n^{(b_0)}, Q_n(x)) < \text{asMSE}_0(\hat{S}_n, Q_n(x))\} \end{aligned} \quad (5.5)$$

**Remark 5.2** (a) The name *cniper* point is due to H. Rieder; it alludes to the fact that this “Janus-type” contamination  $Q_n(x_0)$  pretends to be *nice*, but to the contrary is in fact *pernicious*, “sniping” off the classically optimal procedure. . .

(b) The cniper concept is of course not bound to quadratic loss. In the obvious manor, the concept may be generalized for multivariate observations, if we define any  $x_0$  of minimal absolute as *cniper* point.

(c) To get rid of the dependency upon the radius  $r$ , in the examples we will use the minimax radii  $r_\gamma(n)$  defined in the preceding section.

Correspondingly, in the setup of this paper and under (2.23), let  $S_n^{(c_1)}$  be an M-estimator to the s-o-o IC  $\eta_{c_1}$  according to Corollary 4.2.

**Definition 5.3** *The s-o cniper point  $x_1$  is defined as  $x_{1,+}$  if  $x_{1,+} \geq -x_{1,-}$  and  $x_{1,-}$  else, where*

$$\begin{aligned} x_{1,+} &:= \inf\{x > 0 \mid \text{asMSE}_1(S_n^{(c_1)}, Q_n(x)) < \text{asMSE}_1(\hat{S}_n, Q_n(x))\} \\ x_{1,-} &:= \sup\{x < 0 \mid \text{asMSE}_1(S_n^{(c_1)}, Q_n(x)) < \text{asMSE}_1(\hat{S}_n, Q_n(x))\} \end{aligned} \quad (5.6)$$

*Cniper* contaminations and f/s-o-o ICs form saddle-points under (5.7)/(2.23):

**Proposition 5.4** *The pair  $(S_n^{(b_0)}, Q_n(x_0))$  is a saddlepoint for the class of all pairs  $(S_n, Q_n)$  if*

$$|\hat{\eta}(x_0)| \leq |\eta_b(x_0)| \quad \forall b: \quad |\eta_b(x_0)| < b \quad (5.7)$$

where  $S_n$  are ALE’s to IC’s of form (2.8) and  $Q_n \in \mathcal{Q}_n$  w.r.t. f-o risk  $\tilde{R}$ .

Under (2.23), the same holds in the one-dimensional location model for the pair  $(S_n^{(c_1)}, Q_n(x_1))$  w.r.t. s-o risk in  $\tilde{Q}(r)$ .

**Remark 5.5** A sufficient condition for (5.7) is that  $\Lambda(x) = -\Lambda(-x)$ : Then for any  $b > 0$ ,  $a_b = 0$  is possible and,

$$A_b^{-1} = E \Lambda \Lambda^\tau \min\{1, \frac{b}{|A_b \Lambda|}\} \leq E \Lambda \Lambda^\tau = I$$

So  $A_b \geq I^{-1}$  in the positive semi-definit sense, and hence for  $b$  s.t.  $|\eta_b(x_j)| < b$

$$|\eta_b(x_j)| = |A_b \Lambda(x_j)| \geq |I^{-1} \Lambda(x_j)| = |\hat{\eta}(x_j)| \quad (5.8)$$

### 5.2.2 Error probabilities

For numerical evaluations, we consider the Gaussian location model and the Gaussian location and scale model. In both models,  $x_{j,+} = -x_{j,-}$ , and without loss, we use  $x_{j,+}$ . For the as. tests between  $q_n = p_0$  and  $q_n > p_0$ , alluded to in the beginning of this section, we note that

$$p_0 = P_\theta(X_i \geq x_j) = \Phi(-x_j), \quad q_n = p_0 + \frac{r}{\sqrt{n}}(1 - p_0) \quad (5.9)$$

As to the (f-o) as. minimax test Ruckdeschel (2006, formula (6.1)) gives as as. risk

$$\varepsilon = \varepsilon_\infty = \Phi\left(-\frac{r}{2} \sqrt{\frac{1-p_0}{p_0}}\right) \quad (5.10)$$

For s-o asymptotics, we instead use the finite-sample minimax test, i.e. the Neyman-Pearson test with equal Type-I and Type-II error. In our case this is a corresponding randomized binomial test.

### 5.2.3 Gaussian location

In the Gaussian location model, we draw all necessary expressions from Ruckdeschel (2010a, Prop. ); in particular, with  $c_1 = c_1(n, r_\gamma)$ , and  $A_1 = (2\Phi(c_1) - 1)^{-1}$ ,  $b_1 = c_1 A_1$ , by Theorem 2.1(b), maximizing risk amounts to either  $X^{\text{di}} > c_1$  always or  $X^{\text{di}} < -c_1$  always. The classically optimal estimator is the arithmetic mean, and one easily calculates

$$E_{Q_n(x)}[\bar{x}_n^2 | K = k] = \frac{1}{n^2} [k^2 x^2 + (n - k)] \quad (5.11)$$

and integrating out  $K$  we get directly

$$n E_{Q_n(x)}[\bar{x}_n^2] = 1 - \frac{r}{\sqrt{n}} + x^2(r^2 + \frac{r}{\sqrt{n}} - \frac{r^2}{n}) \quad (5.12)$$

Combining this with formulas (2.17) and (4.1), for  $M_0 := \text{asMSE}_0(S_n^{(c_1)})$  we get

$$x_1^2(n) = \frac{M_0 - 1 + \frac{r}{\sqrt{n}}(M_0 + b_1^2(r^2 + 1) + 1)}{r^2(1 - \frac{1}{n}) + \frac{r}{\sqrt{n}}} \quad (5.13)$$

or

$$x_1(n) = \frac{\sqrt{M_0 - 1}}{r} + \frac{1}{2\sqrt{n}} \left[ \frac{M_0 + 1 + b_1^2(r^2 + 1)}{\sqrt{M_0 - 1}} - \frac{\sqrt{M_0 - 1}}{r^2} \right] + o(\frac{1}{\sqrt{n}}) \quad (5.14)$$

Table 4: Minimax contamination at  $\gamma = 0$ 

$n$	5	10	30	50	100	200	300	$\infty$
$r_\gamma(n)$	0.390	0.449	0.514	0.536	0.559	0.576	0.584	0.621
$c_1(r_\gamma, n)$	0.776	0.749	0.729	0.725	0.722	0.720	0.719	0.718
$x_1(n)$	2.931	2.470	2.101	2.004	1.914	1.853	1.826	1.714
$1 - \beta_n(0.05)$	0.364	0.272	0.215	0.183	0.162	0.133	0.132	0.101
$\varepsilon_n$	0.277	0.178	0.129	0.115	0.097	0.089	0.086	0.072

This yields the results as in Table 4. We include the type-II error  $1 - \beta(\alpha)$  for the Neyman Pearson test to niveau  $\alpha = 5\%$  and the risk  $\varepsilon_n$  of the corresponding minimax test; roughly speaking we cannot do better than overlooking one of 10 contaminations at niveau 5% ideal observations to be falsely marked as outliers, and, equally weighting the two error types we cannot do better than with a false classification rate of 7% for each error type.

#### 5.2.4 Gaussian location and scale

To give one more example, consider the one-dimensional location-scale model at central distribution  $\mathcal{N}(0, 1)$ . For this model we have not yet established a s-o as. theory; for f-o asymptotics, however, we may use R-programs from the bundle RobASt, cf. Kohl (2005, Appendix D), and get  $r_\infty = 0.579$ ,

$$\max_{Q_n \in \mathcal{Q}_n(r_\infty)} \text{asMSE}(\eta_{\theta;0}, Q_n) = 3.123 \quad (5.15)$$

while  $\mathcal{I}_\theta^{-1} \Lambda_\theta = (x, \frac{1}{2}(x^2 - 1))^\tau$ . This gives  $x_0 = 1.844$ —and hence  $\varepsilon_\infty = 5.737\%$  and  $1 - \beta_\infty(5\%) = 6.557\%$ . Condition (5.7) is proved to hold in subsection A.6.

## A Proofs

### A.1 A Hoeffding Bound

**Lemma A.1** Let  $\xi_i \stackrel{\text{i.i.d.}}{\sim} F$ ,  $i = 1, \dots, n$  be real-valued random variables,  $|\xi_i| \leq 1$  Then for  $\mu = \mathbb{E}[\xi_1]$  and  $0 < \varepsilon < 1 - \mu$

$$P\left(\frac{1}{n} \sum_i \xi_i - \mu \geq \varepsilon\right) \leq \left\{ \left( \frac{\mu}{\mu + \varepsilon} \right)^{\mu + \varepsilon} \left( \frac{1 - \mu}{1 - \mu - \varepsilon} \right)^{1 - \mu - \varepsilon} \right\}^n \quad (A.1)$$

*Proof* Hoeffding (1963), Thm. 1, inequality (2.1).  $\square$

To settle case (II) in the proof of Theorem 3.1, we need the following sharpening of Ruckdeschel (2010b, Lem. A.2)

**Lemma A.2** Let  $k_1(n) = 1 + d_n$  and assume that for some  $\delta \in (0, 1/4)$ ,

$$d_n n^{1/4 - \delta} \rightarrow \infty, \quad d_n n^{-1/2 + \delta} \rightarrow 0 \quad \text{for } n \rightarrow \infty \quad (A.2)$$

Then if  $\liminf_n d_n > 0$  there is some  $c > 0$  such that

$$\Pr(\text{Bin}(n, r/\sqrt{n}) > k_1(n)r\sqrt{n}) = o(e^{-cr\sqrt{n}}) \quad (A.3)$$

and, if  $d_n = o(n^0)$ , for any  $0 < \delta_0 \leq 2\delta$ , it holds that

$$\Pr(\text{Bin}(n, r/\sqrt{n}) > k_1(n)r\sqrt{n}) = o(e^{-m^{\delta_0}}) \quad (\text{A.4})$$

**Remark A.3** Even if  $d_n$  is increasing at a faster rate than  $n^{1/2}$ , assertion (A.3) remains true, as long as  $\liminf_n d_n > 0$ —but this is not needed here.

*Proof* Let

$$\mathcal{K}_n := k_1(n) \log k_1(n) + 1 - k_1(n) = \int_1^{k_1(n)} \log(x) dx \quad (\text{A.5})$$

Then  $\mathcal{K}_n > 0$ , as  $\log(x) > 0$  for  $x > 1$  and By the second assumption in (A.2),  $d_n = o(\sqrt{n})$ , so  $0 < d_n r/\sqrt{n} < 1 - r/\sqrt{n}$  eventually in  $n$  and Hoeffding's Lemma A.1 is available; applying it to the case of  $n$  independent  $\text{Bin}(1, p)$  variables, we obtain for  $B_n \sim \text{Bin}(n, p_n)$ ,  $p_n = r/\sqrt{n}$  and  $\varepsilon = (k_1(n) - 1)r/\sqrt{n}$  (which is smaller than  $1 - p_n$  eventually)

$$\begin{aligned} \Pr(B_n > k_1(n)r\sqrt{n}) &\leq \exp\left(-k_1(n)r\sqrt{n} \log(k_1(n)) + (n - k_1(n)r\sqrt{n}) \times \right. \\ &\quad \left. \times (\log(1 - \frac{r}{\sqrt{n}}) - \log(1 - k_1(n)\frac{r}{\sqrt{n}}))\right) \end{aligned}$$

But for  $x_0 < x_1 \in (0, 1)$ ,  $\log(1 - x_0) - \log(1 - x_1) = \int_{1-x_0}^{1-x_1} t^{-1} dt \leq (x_1 - x_0)/(1 - x_1)$ . Thus  $\log(1 - r/\sqrt{n}) - \log(1 - k_1(n)r/\sqrt{n}) \leq \frac{d_n r/\sqrt{n}}{1 - k_1(n)r/\sqrt{n}}$  and

$$\Pr(B_n > k_1(n)r\sqrt{n}) \leq \exp\left(-r\sqrt{n} (k_1(n) \log(k_1(n)) - k_1(n) + 1)\right) = e^{-\mathcal{K}_n r\sqrt{n}},$$

If  $\liminf_n d_n > 0$ , by (A.5)  $\liminf_n \mathcal{K}_n > 0$ , and for any  $0 < c < \liminf_n \mathcal{K}_n$ , (A.3) follows. If  $d_n = o(n^0)$ , we note that

$$\mathcal{K}_n = (1 + d_n) \log(1 + d_n) - d_n = d_n^2/2 + o(d_n^2) \quad (\text{A.6})$$

which for any  $\delta' > 0$  entails

$$\Pr(\text{Bin}(n, r/\sqrt{n}) > k_1(n)r\sqrt{n}) = o\left(\exp\left(-\frac{rd_n^2\sqrt{n}}{2 + \delta'}\right)\right)$$

Now for  $d_n = o(n^0)$ , by the first assumption in (A.2), for  $0 < \delta_0 < 2\delta$  eventually in  $n$ , (A.4) holds as

$$n^{\delta_0} - \frac{d_n^2\sqrt{n}}{2 + \delta'} < n^{2\delta} \left(1 - \frac{n^{1/2-2\delta}d_n^2}{2 + \delta'}\right) \rightarrow -\infty$$

□

Another consequence of the exponential decay of (A.3)/(A.4) is that we may neglect values of  $K > k_1(n)r\sqrt{n}$  when integrating along  $K$ .

**Corollary A.4** Let  $K \sim \text{Bin}(n, r/\sqrt{n})$ . Then, in the setup of Lemma A.2, for any  $j \in \mathbb{N}$ ,

$$\mathbb{E}[K^j \mathbf{I}_{\{K \geq k_1(n)r\sqrt{n}\}}] = o(e^{-m^d}) \quad (\text{A.7})$$

for any  $0 < d < \sqrt{n}$  if  $\liminf_n d_n > 0$  and any  $0 < d \leq \delta_0$  if  $\lim_n d_n = 0$ .

*Proof*  $\mathbb{E}[K^j \mathbf{I}_{\{K \geq k_1(n)r\sqrt{n}\}}] \leq n^j \Pr(K > k_1(n)r\sqrt{n}) \stackrel{(\text{A.3})/(\text{A.4})}{=} o(e^{-m^d})$ . □

## A.2 Proof of Theorem 3.1

In the risk, we have to treat stochastic arguments in  $\Phi, \varphi$ ; this is settled in the following lemma:

**Lemma A.5** *Let  $F: \mathbb{R} \rightarrow \mathbb{R}$  be twice differentiable with Hölder-continuous second derivative and  $G: \mathbb{R} \rightarrow \mathbb{R}$  be differentiable with Hölder-continuous derivative. Then there is a sequence  $k_1(n) = 1 + d_n$  with  $d_n \rightarrow 0$  according to (A.2) and some  $\eta > 0$ , such that for all  $x, \beta \in \mathbb{R}$  and with  $\tilde{k} = K/\sqrt{n}$ ,*

$$\mathbb{E}[F(x + \beta\tilde{k}) | K \leq k_1(n)r\sqrt{n}] = F(x + \beta r) + F''(x + \beta r) \frac{\beta^2 r}{2\sqrt{n}} + o(n^{-1/2}) \quad (\text{A.8})$$

and

$$\mathbb{E}[G(x + \beta\tilde{k}) | K \leq k_1(n)r\sqrt{n}] = G(x + \beta r) + O(n^{-(1+\eta)/4}) \quad (\text{A.9})$$

*Proof* Using the Taylor approximation of  $\log(1+x)$ , we get for  $n$  sufficiently large

$$d_n^2/3 \leq d_n^2/2 - d_n^3/6 \leq \mathcal{K}_n \leq d_n^2/2 \quad (\text{A.10})$$

By (A.4) of Lemma A.2, for some  $\delta_0$  and eventually in  $n$  we have  $P(K > k_1(n)r\sqrt{n}) \leq \exp(-rn^{\delta_0})$ , and by the same argument we also get that  $P(K < (2 - k_1(n))r\sqrt{n}) \leq \exp(-rn^{\delta_0})$ . Hence,

$$P(|\tilde{k} - r| > rd_n) \leq 2 \exp(-rn^{\delta_0}) \quad (\text{A.11})$$

Thus, as  $F, G$  are bounded, the contribution of the set  $\{|\tilde{k} - r| > rd_n\}$  decays exponentially, while on the complement we have a uniformly bounded Taylor expansion up to order 2 respectively 1 for the integrands:

$$\begin{aligned} F(x + \beta\tilde{k}) &= F(x + \beta r) + F'(x + \beta r)\beta(\tilde{k} - r) + F''(x + \beta r)\beta^2(\tilde{k} - r)^2/2 + o((\tilde{k} - r)^{2+\eta}) \\ G(x + \beta\tilde{k}) &= G(x + \beta r) + G'(x + \beta r)\beta(\tilde{k} - r) + o((\tilde{k} - r)^{1+\eta}) \end{aligned}$$

Integrating these expansions out in  $\tilde{k}$ , we see that the first contribution to the Taylor series for  $F$  is the quadratic term, which is  $F''(x + \beta r) \frac{\beta^2}{2} \text{Var } \tilde{k}$ , and the remainder is  $o(n^{-1/2})$ . For  $G$ , the first contribution to the error term is the remainder, hence of form  $\text{const}|\tilde{k} - r|^{1+\eta}$ . By the Hölder inequality this gives a bound  $\text{const}[\text{Var } \tilde{k}]^{\frac{1+\eta}{2}} = O(n^{-(1+\eta)/4})$ .  $\square$

For the proof of Theorem 3.1, we use a tableau like the one of Ruckdeschel (2010a, p. 19), i.e., to derive the result, we partition the integrand according to

	$K < k_1(n)r\sqrt{n}$	$k_1(n)r\sqrt{n} \leq K < \varepsilon_0 n$
$ t  \leq k_2 b^2 \log(n)/n$	(I)	(II)
$k_2 b^2 \log(n)/n <  t $	(III)	

with  $k_1(n)$  according to (A.2). This time, no integration w.r.t.  $t$  is needed, so case (IV) from Ruckdeschel (2010a) may be canceled, which is why we may dispense of assumption (Pd) and pass to the unrestricted neighborhoods  $\mathcal{Q}_n$ . Cases (II) and (III) may be taken over unchanged from Ruckdeschel (2010a, Proof of Thm. 3.5), so we may confine us to case (I):

We use  $\alpha_1, \alpha_2$  from (3.2) and proceed paralleling the proof in Ruckdeschel (2010a) and get from formula (A.18) therein that  $\Pr(S_n \leq -\frac{\alpha_1}{\sqrt{n}} | D_{k,\tilde{t}}) = \tilde{G}_n(-\frac{\alpha_1}{\sqrt{n}}) + O(n^{-3/2})$ . So we have to spell out  $s_{n,k}(-\frac{\alpha_1}{\sqrt{n}})$ , which gives

$$s_{n,k}(-\frac{\alpha_1}{\sqrt{n}}) = v_0^{-1} \left\{ (-t^{\sharp} - \alpha_1) + \frac{1}{\sqrt{n}} \left[ \frac{\tilde{k}}{2} \alpha_1 - \alpha_1 \tilde{v}_1(t^{\sharp} + \alpha_1) - \frac{l_2}{2} \alpha_1^2 \right] \right\} + o(\frac{1}{\sqrt{n}}) \quad (\text{A.12})$$

and hence—setting  $\tilde{s} = s_{n,k}(-\frac{\alpha_1}{\sqrt{n}})$  and  $\tilde{s}_1 = -(\alpha_1 + t^{\sharp})/v_0$  as in Ruckdeschel (2010a)

$$\begin{aligned} \Pr(S_n \leq -\frac{\alpha_1}{\sqrt{n}} | D_{k,\tilde{t}}) &= \Phi(\tilde{s}) - \varphi(\tilde{s}) \frac{(\tilde{s}^2 - 1)}{6\sqrt{n}} \rho(-\frac{\alpha_1}{\sqrt{n}}) + o(\frac{1}{\sqrt{n}}) = \\ &= \Phi(\tilde{s}_1) + \frac{\varphi(\tilde{s}_1)}{2\sqrt{nv_0}} [\alpha_1 \tilde{k} - l_2 \alpha_1^2 - 2(\alpha_1 + t^{\sharp}) \tilde{v}_1 \alpha_1 - v_0 \frac{\rho_0}{3} (\tilde{s}_1^2 - 1)] + o(\frac{1}{\sqrt{n}}) \quad (\text{A.13}) \end{aligned}$$

This term is maximized eventually in  $n$ , if  $-t^{\sharp}$  is maximal or, essentially equivalent, all contaminating mass (up to mass  $\mathbf{o}(n^{-1/2})$ ) is concentrated left of  $\tilde{y}_n$  from Section 2.2, and then  $t^{\sharp} = k^{\sharp}\tilde{b}$ , and after the substitution according to  $\tilde{k} := k/\sqrt{n}$ ,  $k^{\sharp} := k/\sqrt{n}$ , this gives with  $\tilde{s}_k = -(\alpha_1 + \tilde{k}\tilde{b})/v_0$

$$\Pr(S_n \leq -\frac{\alpha_1}{\sqrt{n}} | D_{k,\tilde{k}=\tilde{k}\tilde{b}}) = \Phi(\tilde{s}_k) + \frac{\varphi(\tilde{s}_k)}{2\sqrt{nv_0}} [\alpha_1 \tilde{k} - l_2 \alpha_1^2 - 2\tilde{s}_k v_0 \tilde{v}_1 \alpha_1 - v_0 \frac{\rho_0}{3} (\tilde{s}_k^2 - 1) - \tilde{k}^2 \tilde{b}] + \mathbf{o}(\frac{1}{\sqrt{n}}) \quad (\text{A.14})$$

Now, by (3.6), it holds that  $s_1 = -(\alpha_1 + r\tilde{b})/v_0$ , so that by an application of Lemma A.5, for  $Q_{n;-}^0$  any sequence of measures according to (2.20)

$$Q_{n;-}^0(S_n \leq -\frac{\alpha_1}{\sqrt{n}}) = \Phi(s_1) + \mathbf{o}(\frac{1}{\sqrt{n}}) + \frac{1}{\sqrt{n}} \varphi(s_1) \left[ \frac{r}{2v_0} \alpha_1 - \frac{l_2}{2v_0} \alpha_1^2 + s_1 v_0 \tilde{v}_1 \alpha_1 - \frac{\rho_0}{6} (\tilde{s}_1^2 - 1) - r \frac{\tilde{b}^2}{2v_0^2} s_1 - r^2 \frac{\tilde{b}}{2v_0} \right]$$

Correspondingly, we get for any sequence of measures  $Q_n^+$  according to (2.21)

$$Q_{n;+}^0(S_n \geq \frac{\alpha_2}{\sqrt{n}}) = \Phi(s_1) + \mathbf{o}(\frac{1}{\sqrt{n}}) + \frac{1}{\sqrt{n}} \varphi(s_1) \left[ \frac{r}{2v_0} \alpha_2 + \frac{l_2}{2v_0} \alpha_2^2 - s_1 v_0 \tilde{v}_1 \alpha_2 + \frac{\rho_0}{6} (\tilde{s}_1^2 - 1) - r \frac{\tilde{b}^2}{2v_0^2} s_1 + r^2 \frac{\tilde{b}}{2v_0} \right]$$

We next account for order  $\frac{1}{\sqrt{n}}$ -terms and get, as  $\delta' = \mathbf{O}(\frac{1}{\sqrt{n}})$

$$Q_{n;-}^0(S_n \leq -\frac{\alpha'_1}{\sqrt{n}}) = Q_{n;-}^0(S_n \leq -\frac{\alpha_1}{\sqrt{n}}) + \delta' \varphi(\frac{a-r\tilde{b}}{v_0}) + \mathbf{o}(\frac{1}{\sqrt{n}}) \quad (\text{A.15})$$

and analogously for  $Q_{n;+}^0(S_n \geq \frac{\alpha'_2}{\sqrt{n}})$ , so  $\delta' = \frac{1}{\sqrt{n}} \left( -\frac{r\delta}{2v_0} - \frac{l_2}{2v_0} (a^2 + \delta^2) - \tilde{v}_1 v_0 s_1 \delta - \frac{\rho_0}{6} (s_1^2 - 1) + \frac{r\tilde{b}\delta s_1}{v_0^2} + \frac{r^2 \delta}{2v_0} \right)$

and  $Q_{n;-}^0(S_n \leq -\frac{\alpha'_1}{\sqrt{n}}) = Q_{n;+}^0(S_n \geq \frac{\alpha'_2}{\sqrt{n}}) + \mathbf{o}(\frac{1}{\sqrt{n}})$ , i.e.,

$$Q_{n;-}^0(S_n \leq -\frac{\alpha'_1}{\sqrt{n}}) = \Phi(s_1) + \varphi(s_1) \frac{1}{\sqrt{n}} \left[ \frac{ra}{2v_0} + 2 \frac{l_2 a \delta}{v_0} - a s_1 \tilde{v}_1 - \frac{r(b^2 + \delta^2) s_1}{4v_0^2} + \frac{r^2 \tilde{b}}{2v_0} \right] + \mathbf{o}(\frac{1}{\sqrt{n}}) \quad (\text{A.16})$$

□

### A.3 Proof of Corollary 4.2

The assumptions of Theorem 4.1 are clearly fulfilled. Hence we may start with the verification (4.3):

$$G(w, s) = (w^2 + s^2) \left( 1 + \frac{r}{\sqrt{n}} \right) + \frac{r}{\sqrt{n}} w^2 \left( 1 + \frac{1}{r^2} \right) \quad (\text{A.17})$$

$$\partial_w G(w, s) = 2w \left[ 1 + \frac{r}{\sqrt{n}} + \frac{r}{\sqrt{n}} \left( 1 + \frac{1}{r^2} \right) \right], \quad \partial_s G(w, s) = 2s \left[ 1 + \frac{r}{\sqrt{n}} \right] \quad (\text{A.18})$$

and hence, dividing both sides of (4.2) by  $2\hat{A}\hat{v}_0$ , we get the assertion. The LHS of (4.3) (with or without factor  $1 + \frac{r^2+1}{r^2+r\sqrt{n}}$ ) is isotone, the RHS antitone in  $c$ . Thus if we insert the factor to correct the f-o-o clipping height  $c_0$  to  $c_1(n)$ , the factor increases the LHS without affecting the RHS. This can only be compensated for by a decrease of  $c_0$  to  $c_1(n)$ . If  $h(c)$  is differentiable in  $c_0$  with derivative  $h'(c_0)$ , (4.4) is an application of the applying the implicit function theorem: Let  $G(s, c) := r^2 c(1 + s) - h(c)$ . Then  $G(0, c_0) = 0$ . Hence for  $s = (r^2 + 1)/(r^2 + r\sqrt{n})$ , up to  $\mathbf{o}(n^{-1/2})$ ,

$$c_1(n) + \mathbf{o}(n^{-1/2}) = c_0 - \frac{G_s(0, c_0)}{G_c(0, c_0)} s = c_0 \left( 1 - \frac{1}{\sqrt{n}} \frac{r^3 + r}{r^2 - h'(c_0)} \right) + \mathbf{o}(n^{-1/2})$$

□

#### A.4 Proof of Proposition 4.3

We apply Rieder (1994, Theorem 1.4.7) to the derivatives; this theorem says that for  $\eta \in C_1(\mathbb{R})$  with  $\eta(\theta_0) = 0$ ,  $\eta'(\theta_0) \neq 0$  for some  $\theta_0 \in \mathbb{R}$ , there exists an open neighborhood  $V_0 \subset C_1(\mathbb{R})$  such that for every open, connected neighborhood  $V \subset V_0$  of  $\eta$  there is a unique, continuous map  $T: V \rightarrow \mathbb{R}$  with

$$T(\eta) = \theta_0, \quad f(T(f)) = 0, \quad f \in V \quad (\text{A.19})$$

even more so,  $T$  is continuously bounded differentiable on  $V$  with derivative at tangent  $h$

$$dT(f)h = -h(T(f))/f'(T(f)) \quad (\text{A.20})$$

Hence there is an open neighborhood  $V_{0,F}$  of  $F$  such that for each connected open neighborhood  $V_F \subset V_{0,F}$ , we get a unique, continuously bounded differentiable map  $T: V_F \rightarrow \mathbb{R}$  with

$$T(F) = x_0, \quad f'(T(f)) = 0, \quad f \in V_F, \quad dT(f)h = -h'(T(f))/f''(T(f)) \quad (\text{A.21})$$

But by assumption (4.5) from some  $n$  on,  $F_n$  and  $G_n$  will lie in  $V_{0,F}$ , and setting  $x_n = T(F_n)$ , by (A.21)  $F_n'(x_n) = 0$ , and

$$|x_n - x_0| = |T(F_n) - T(F)| \leq |F_n'(x_0)|/F''(x_0) = O(n^{-\beta'})$$

which is (b); again by (4.5),

$$|F_n''(x_n) - F''(x_0)| \leq |F_n''(x_n) - F''(x_n)| + |F''(x_n) - F''(x_0)| \leq \sup_x |F_n''(x) - F''(x)| + o(n^0) = o(n^0)$$

In particular, eventually in  $n$ ,  $F_n''(x_n) > 0$  and hence  $x_n$  is a minimum of  $F$ , so (a) is shown. By (4.5),  $\sup_x |F - G_n| + |F' - G_n'| + |F'' - G_n''| = O(n^{-\beta'})$ , so (c) follows just as (a). For (d) we note

$$|x_n - y_n| = |T(F_n) - T(G_n)| \leq |G_n'(x_n)|/F_n''(x_n) \stackrel{(a)}{=} |G_n'(x_n)|/(f_2 + o(n^0)) = O(n^{-\beta})$$

To show (e), we introduce  $d_n := y_n - x_n$  and write

$$0 \leq G_n(x_n) - G_n(y_n) = G_n'(y_n)d_n + G_n''(y_n)d_n^2/2 + o(d_n^2) = (f_2 + o(n^0))d_n^2/2 + o(d_n^2) = O(n^{-2\beta}) \quad (\text{A.22})$$

□

#### A.5 Proof of Proposition 5.4

We show that under the assumptions of this proposition  $x_j$  indeed defines a “uniformly bad contamination” in the sense that for the fixed contamination  $Q_n(x_j)$

$$\text{asMSE}_0(S_n^{(b_0)}, Q_n(x_0)) = \min_{b>0} \text{asMSE}_0(S_n^{(b)}, Q_n(x_0)) \quad (\text{A.23})$$

resp.  $\text{asMSE}_1(S_n^{(c_1)}, Q_n(x_1)) = \min_{c>0} \text{asMSE}_1(S_n^{(c)}, Q_n(x_1))$  In case  $j = 0$ , as in the setup of Rieder (1994, chap. 5), we obtain

$$\text{asMSE}_0(S_n^{(b)}, Q_n(x_0)) = \text{tr Cov}_{\text{id}}(\eta_b) + r^2 |\eta_b(x_0)|^2, \quad \text{asMSE}_0(\hat{S}_n, Q_n(x_0)) = \text{tr } \mathcal{I} + r^2 |\hat{\eta}(x_0)|^2 \quad (\text{A.24})$$

Now for given  $x_0$ , either  $|\eta^{(b)}(x_0)| < b$  or  $|\eta^{(b)}(x_0)| = b$ . In the first case, (5.7) applies and hence

$$\text{asMSE}_0(S_n^{(b_0)}, Q_n(x_0)) \geq \text{asMSE}_0(\hat{S}_n, Q_n(x_0)) \quad (\text{A.25})$$

In the latter,  $Q_n(x_0)$  already achieves maximal as. risk for  $S_n^{(b)}$  on  $Q_n$ , and hence by minimaxity of  $S_n^{(b_0)}$

$$\text{asMSE}_0(S_n^{(b)}, Q_n(x_0)) \geq \text{asMSE}_0(S_n^{(b_0)}, Q_n(x_0)) \quad (\text{A.26})$$

For the case  $j = 1$  one argues in an analogue way. □

## A.6 Proof for (5.7) in the Gaussian location scale model

We abbreviate the location and scale parts by indices  $l$  and  $s$  respectively. By equivariance we may limit ourselves to the case  $\theta = (0, 1)^T$ . Due to symmetry,  $A = A(b)$  from (2.8) is diagonal for all  $b$  with elements  $A_l$  and  $A_s$  and we may write

$$\eta_b = Y \min\{1, b/|Y|\}, \quad Y^T = (A_l x, A_s(x^2 - 1 - z_s)) \quad (\text{A.27})$$

The centering  $z_s(b)$  after the clipping is necessary, as the scale part is not skew symmetric; in the pure scale case (with known  $\theta_l$ ), the corresponding centering  $z'_s = z'_s(b)$  is antitone in  $b$ , because  $A_s$  is monotone in  $x^2$ : It decreases from 0 to  $[\Phi^{-1}(3/4)]^2 - 1 \doteq -0.545 =: \check{z}$ . In the combined case, we never reach this extremal case due to the additional location part—compare Kohl (2005, Remark 8.2.1(a)) where  $\bar{z}_s = \bar{a}_{sc}/\bar{\alpha} - 1 \doteq -0.530$ ; in any case,  $z_s > -1$  always. Hence in particular, for  $x_0 = 1.844$  and  $b$  such that  $|\eta^{(b)}(x_0)| \leq b$  it holds that

$$|\eta_s^{(b)}(x_0)| = A_s(b)|x_0^2 - 1 - z_s(b)| > A_s(b)|x_0^2 - 1| > I_s^{-1}|x_0^2 - 1| = |\hat{\eta}_s(x_0)| \quad (\text{A.28})$$

and thus in particular,

$$|\eta^{(b)}(x_0)|^2 = |\eta_s^{(b)}(x_0)|^2 + |\eta_l^{(b)}(x_0)|^2 = |\eta_s^{(b)}(x_0)|^2 + A_{0,l}(b)x_0^2 > \hat{\eta}_s(x_0)^2 + I_l^{-2}x_0^2 = |\hat{\eta}(x_0)|^2 \quad (\text{A.29})$$

□

## Acknowledgement

## References

- Andrews D.F., Bickel P.J., Hampel F.R., Huber P.J., Rogers W.H. and Tukey J.W. (1972): *Robust estimates of location. Survey and advances*. Princeton University Press, Princeton, N. J.
- Donoho D.L. and Huber P.J. (1983): The notion of breakdown point. In: Bickel, P.J., Doksum, K.A. and Hodges, J.L.jun. (eds.) *Festschr. for Erich L. Lehmann*, pp. 157–184.
- Fraiman R., Yohai V.J. and Zamar R.H. (2001): Optimal robust  $M$ -estimates of location. *Ann. Stat.*, **29**(1): 194–223.
- Hoeffding W. (1963): Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.*, **58**: 13–30.
- Huber P.J. (1968): Robust confidence limits. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **10**: 269–278.
- (1981): *Robust statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley.
- (1997): *Robust statistical procedures*, Vol. 68 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2. edition.
- Kohl M. (2005): *Numerical contributions to the asymptotic theory of robustness*. Dissertation, Universität Bayreuth, Bayreuth.
- Kohl M., Ruckdeschel P. and Rieder H. (2010): Infinitesimally robust estimation in general smoothly parametrized models. *Statistical Methods and Applications*. To appear. DOI: [10.1007/s10260-010-0133-0](https://doi.org/10.1007/s10260-010-0133-0).
- Pfanzagl J. (1979): First order efficiency implies second order efficiency. In: *Contributions to statistics, Jaroslav Hajek Mem. Vol.*, 167–196.
- R Development Core Team (2010): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. **URL:** <http://www.R-project.org>
- Rieder H. (1980): Estimates derived from robust tests. *Ann. Stat.*, **8**: 106–115.
- (1994): *Robust asymptotic statistics*. Springer Series in Statistics. Springer.
- Rieder H., Kohl M. and Ruckdeschel P. (2008): The costs of not knowing the radius. *Statistical Methods and Applications*, **17**(1): 13–40.
- Ruckdeschel P. (2006): A motivation for  $1/\sqrt{n}$ -Shrinking-Neighborhoods? *Metrika*, **63**(3): 295–307.
- (2010a): Higher Order Asymptotics for the MSE of  $M$ -Estimators on Shrinking Neighborhoods. Submitted; ArXiv Nr. .
- (2010b): Higher Order Asymptotics for the MSE of the Median on Shrinking Neighborhoods. Submitted; ArXiv Nr. .



- 
- Ruckdeschel P. and Kohl M. (2010): Computation of the Finite Sample Risk of M-Estimators on Neighborhoods. Submitted; available on arXiv:.
- Ruckdeschel P. and Rieder H. (2004): Optimal influence curves for general loss functions. *Stat. Decis.*, **22**: 201–223.
- 

Web-page to this article:

<http://www.mathematik.uni-kl.de/~ruckdesc/>